DOCUMENT RESUME

ED 427 076 TM 029 457

AUTHOR Dawson, Verdell Lett; Acker-Hocevar, Michele
TITLE Contradictions within a State-Mandated Teacher Evaluation System: A Historical Perspective, Personal Reflections and Principal Interviews.
PUB DATE 1998-11-00
NOTE 28p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (New Orleans, LA, November 4-6, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Assistant Principals; Case Studies; Educational Change; *Educational History; Elementary Secondary Education; Evaluation Methods; *Principals; Program Implementation; School Districts; *State Programs; *Teacher Evaluation

ABSTRACT

The district-level implementation of a systemic educational change in teacher evaluation mandated at the state level was studied at the elementary and middle school level in a small rural school system. The historical background of teacher evaluation processes from early American schooling to the 1909s was studied. The pilot case study then examined some personal reflections and the experiences and findings of seven principals and five assistant principals. The historical exploration demonstrated a movement from checklists of character traits toward a system that focused on both improved teaching and improved student learning. Personal reflections of one principal suggested that the present evaluation system was a positive reform. Results from the administrators show that paradoxes and contradictions exist inherently within teacher evaluation. The multiple data sources show an awkward and complex process of evaluation. However, the difficult and awkward process results in a more accurate and fair method of assessing how teachers actually affect student learning. The current teacher evaluation system, although imperfect and cumbersome, offers a better process than previously used methods. (Contains 27 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 427 076

Contradictions within a State-Mandated Teacher Evaluation System: A Historical Perspective, Personal Reflections and Principal Interviews

Verdell Lett Dawson
Doctoral Student
The University of Alabama
And
Michele Acker-Hocevar, Ph. D.
Assistant Professor
The University of Alabama

P. O. Box 870231
Administration and Educational Leadership
College of Education
Tuscaloosa, Alabama 35487-0231
(205) 348-1161
Macker@bamaed.ua.edu

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Verdell Dawson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the 1998 Annual Mid-South Educational Research Conference
New Orleans, Louisiana

TM029457

BEST COPY AVAILABLE

Abstract

The pilot case study examined (a) the historical background of teacher evaluation processes from early American schooling to the 1990s, (b) personal reflections and (c) principals' experiences and findings. Supreme questions for the study centered upon local implementation of a state-mandated teacher evaluation system within the elementary and middle school classrooms of a small rural school system. The pilot case study's ultimate purpose was to determine whether principals at the grass roots level would adhere to the beliefs and practices of the State's valid and reliable research-based system. Simply stated, would extensively trained principals follow the State plan was the overriding question.

A historical perspective demonstrated a movement from checklists of character traits toward a system that focused on both improved teaching and improved student learning. Personal reflections revealed that my own individual experience as a teacher, a principal, a supervisor of teachers and a district coordinator for teacher evaluation suggested that the present evaluation system was a positive reform.

The pilot case study revealed that paradoxes/contradictions inherently existed within teacher evaluation. Multi-data sources disclosed an awkward, complex process. Yet, the difficult, awkward process resulted in a more accurate and fair method of assessing how teachers actually effect student learning.

The summary of data sources: (a) a historical perspective, (b) personal reflections, (c) principal interviews was conclusive. Similar themes emerged despite the variation in instrumentation or data collection. Although imperfect and cumbersome, the reviewed teacher evaluation system offered a better process than previously utilized methods, thus offering hope as one answer to the age-old problem of assessing teacher performance.

Introduction

Educational accountability is a continuous expectation of the supporters of public education. At times, the outcry for quality education is loud; at other times, it is audible, but soft. Yet, at almost no time is it inaudible. Much of this push for accountability centers upon the evaluation of teachers. Seemingly, American society believes that schools are only as good as are their teachers. Schools use various methods to evaluate teachers. Within the education community the common use of some of these methods cause them to be identified as measures of past practice, meaning they are accepted by the profession without alarm or concern.

Educational reformers, though, have the responsibility of improving the performance of schools. Sometimes this means studying current and past beliefs and practices of schools and their influencing factors. The result is change. By nature, humans are apprehensive about change. When change relates to their job or livelihood, teachers, like other people, become frightened.

This pilot case study seeks to examine systemic change mandated at the state level, and implemented by numerous local personnel at the district level. The aforementioned is the global scope of the problem. The restricted scope is that of a small school system in rural Ransom.

Background

Teacher evaluation is cyclical, meaning that as a topic of reform, teacher evaluation is more critical at some times than at other times. In the State of Ransom, educational accountability is an exigent issue. Elected officials, want-to-be elected officials, business people, industrialists, chambers of commerce personnel, tourism personnel, retirees, parents and guardians are all placing the current and future economic progress of Ransom at the schoolhouse and classroom doors. Thus, the Ransom State Board of Education and the Ransom Legislature, in joint, but separate actions, mandated and made legal the evaluation of all professional personnel. The Ransom State Department of Education advanced this cry from the voting public by studying, designing and developing an evaluation system for the separate categories of professional education

personnel. The system of evaluation is titled "Certificated Education Personnel Evaluation Program of Ransom, (CEPE)."

Ransom began development of its CEPE evaluation system during the 1980s. During the early days of development, the Ransom State Superintendent invited local boards of education to develop their own systems. To assist them, from December 1989 to September 1990 personnel from all systems trained two-five days per month. Ransom Department of Education personnel and consultants-researchers who were university personnel provided training. After the training local districts that chose to develop their own evaluation system(s) proceeded. Before implementation, however, developed systems were to be submitted to the Ransom Department of Education for approval. Many systems, especially smaller ones with limited resources, did not meet Ransom Department of Education standards. Ultimately, ninety-nine percent of Ransom's 154 local boards of education chose to accept the state-developed model for the evaluation of certificated personnel (Evaluator Manual, 1997)

As of the 1998-99 scholastic year, the Ransom Department of Education designed and generated an evaluation system for five categories of certificated educational personnel: 1) superintendent; 2) central office personnel (assistant superintendent, director, supervisor and program specialist); 3) principals; 4) assistant principals; and 5) teachers. The administrator system-- field tested statewide and implemented for a period of five years or more—was the first of the evaluation systems to be developed. In the early days of implementation, local boards of education chose to use or not to use the state system to evaluate administrators, especially superintendents. The administrator system and the teacher system began in the waning days of the former State Superintendent. He was State Superintendent for several decades. With the advent of the new State Superintendent of Education, CEPE dismantled. Reorganization of the responsible division brought about retirements and new assignments/transfers.

One year later, CEPE was revitalized. Its previous version for teachers was a formative system. The new and revised version had a possible added feature; local boards of education could adopt a summative component. The statewide mandatory beginning date for implementation of all administrator systems in Ransom's 154 school systems was 1996-97. The mandatory starting date for the teacher system was 1997-98. CEPE

officially recognizes that year as 'the start-up year.' This researcher is studying the implementation of CEPE: The Teacher System in the Anderson School System, Anderson, Ransom. The title of the project is "Contradictions within a State-mandated Teacher Evaluation System: A Historical Perspective, Personal Reflections and Principal Interviews."

PROBLEM

Anderson, Ransom, a quaint historical town with a population of 40,000, chronicles its past, from its role as the second biggest manufacturing center during the River Wars, to the Rights Movement. Historians assert that many events in this seemingly sleepy little town changed the course world civilization. Reflective of its River Wars history, visitors from around the nation and the world delight in visiting the picturesque, historically accurate period homes. They also marvel at the variety of historical and cultural sites that center on regional and national museums. Anderson offers a serene bucolic location in the midst of thousands of acres of forest land, and is centrally located within the State, only 50 miles from the state capital, and 175 miles from the state's leading industrial city.

The Anderson School System is composed of twelve elementary schools serving grades kindergarten through six, a prekindergarten program for four-year-old which operates at eight elementary schools, two middle schools for grades seven and eight, and one high school which serves grades nine through twelve. There are 8,500 students enrolled in these schools. Approximately 95 % of them are African-American. No less than 70 % of students are eligible for participation in the free and reduced lunch program.

The Anderson School System has a commitment to excellence converging on providing effective instructional leadership, responsible fiscal management, and quality learning environments for students. Beginning with the 1995-1996 school year, the Anderson School System implemented an uniform dress code for grades kindergarten through twelve that has resulted in higher morale, more orderly classroom behavior, and an end to the 'fashion wars' that had traditionally distracted students from classroom work.

Teachers in the Anderson City School System embraced the practice that principals would visit classrooms once per semester or twice per year for the purpose of

teacher evaluation. Policy of the Anderson Board of Education established this practice. Principals or assistant principals would announce their visit for the first observation, but would keep the date of the second visit secret, surprising teachers with their arrival. Because of the announced or planned status of visits, teachers and principals expected the first observation to be a demonstration of the **best** teaching performance each teacher could display. Before this announced visit, principals would speak briefly or extensively (personal choice) to teachers to determine the objectives and activities of the class time they would observe.

Next, for both the announced and the unannounced observation, teachers expected principals to write or record a full narrative account of every action, statement, and environmental factor of the learning environment. After principals left the classroom, they would assess strengths and weaknesses of the observed learning situation. Unveiling strengths and weaknesses would be the purpose of the post observation conference that followed each observation.

Near the end of the second semester, principals and teachers in the evaluation cycle met to discuss the annual evaluation report. The annual report provided a rating of teachers in three areas: instruction, management and professional responsibilities. After an end-of-the-year conference, the annual report was forwarded to the Superintendent and Board of Education, each of whom would review the report for job decisions before placement in the personnel file of the individual teacher. The Knowledge Evaluation System was the officially adopted evaluation system. Its use was comfortable, established practice.

The outcomes from the annual report were the basis of the development of the individual Professional Development Plan. Teachers were expected to improve in two identified areas of weakness. Attendance at workshops and the reading of professional periodicals as well as other activities that accumulated continuing equivalency units (CEUs) needed for teacher certification would make up the agreed upon professional development plan. Agreeing to the plan were the principal and the teacher along with the Superintendent or her designee giving the final approval or disapproval.

Teachers and principals knew what to expect of the evaluation system. Both teachers and principals understood the process. Teacher knew that every three years,

they would be evaluated individually according to a multi-year cycle. The overall rating would include some high scores and some low scores.

All administrators (superintendent, central office personnel, principals and assistant principals) of Anderson School System are certified evaluators of the Ransom Administrator System. They trained for ten days in eight-hour sessions over the course of one school year. They successfully passed several reliability tests of performance in scripting and a knowledge test with a minimum of 80 % mastery.

The Superintendent uses this System to evaluate administrators annually. Their knowledge of this evaluation system aided them in preparation for their personal annual review. Additionally, principals and assistant principals were certified evaluators of the original CEPE system for teachers. Ten additional days of eight-hour sessions with instruction, reliability tests and a knowledge test with a minimum of 80 % mastery earned them their new certification.

Historical Perspective

Teacher evaluation systems traditionally are designed to serve two purposes: accountability or the measurement of teacher competence and professional development and growth (Boyd, 1989; Lofton, Hill & Claudet, 1997). Educational accountability programs that measure school progress are institutions of contemporary society. In 1983 thirty-eight states mandated some form of accountability legislation (Keefover, 1983). Yet, says M. D. Thomas (1979), studies of the history of education point to the omission of teacher evaluation from the evaluation of schools. Thomas (1979) further asserts that areas studied were architecture, expenditures per pupil, curriculum guides, number of books in the library, course selection lists, number of certified staff per thousand students, administrative organization and degrees obtained by staff.

Researchers, theorists and practitioners alike agree that performance evaluation of teachers is the key to quality or excellence in education. "Schools will be accountable when individual teacher performance is accountable" (Thomas, 1979, p. 9). Even in empirical literature, teacher evaluation as a researched field is veiled. Various voices decry poor student performance and the politicizing of educational accountability. M. D. Thomas, a superintendent-practitioner who has had a long time interest in performance evaluation of educational personnel, is one voice that addresses this subject in a most

direct and comprehensive manner. His research then is the primary basis for many orientations and subsequent findings of this pilot case study. In addition, his research provides a guide to the thinking of theorists before and during the development of the evaluation system on which this pilot case study focuses.

"Teacher Characteristics" Method

Historically, educators were evaluated by character traits (Manning, 1988; Thomas, 1979). The quality teacher was one who had "good grooming" or "good speech" or "good looks" or the "good, pleasant personality." Such traits became items in a checklist that resulted in teachers rated "good" or "bad." Doing the ratings were "superiors considered to possess these traits to a very high degree" themselves, enabling them to "find" and "judge" or "rate" these characteristics in others (Thomas, 1979).

Checklists that measure degrees of accomplishment, according to Manning (1988), are now obsolete. Yet, devised more than eighty years ago, and despite the researched revision of desirable teacher characteristics, the checklist system of today is a close match to that of yesteryear. It is surprising that this instrument does not reflect the latest research findings on teacher evaluation. Manning (1988) suggests that the use of old methods such as checklists will not advance the learning and use of effective teaching practices. In fact, they might impede them.

Traits listed on the traditional checklist usually included enthusiasm, strong voice, flexibility, humor, personality, insight, judgment, originality, adaptability, initiative, persistence, self-confidence, integrity, conviction, cooperation, leadership, punctuality, pleasant appearance, good diction and morality (Thomas, 1979). When studying individual school systems, Thomas (1979) found one school system listed "sincerity," self-control," "alertness," "pleasant mannerism," and "good grooming" as a measure for evaluating teachers. A second school system listed "attendance at professional meeting," "reading" of professional journals," "acceptance of majority decisions," and "compliance with school system decision." (Thomas, 1979).

Education research has not connected the quality of performance to character traits. The continued use of checklists of character traits reinforces the position of Robert Finley, superintendent of schools in Glen Cove, New York during the Seventies, "

Evaluation is subjective...period. No other way to evaluate people exists—so that's the way to do it." (Thomas, 1979, p. 11).

The trend in teacher evaluation shifted when people began to question the use of traits in assessment. They acknowledged that educators like other workers share high degrees of intelligence, emotional stability, energy level and motivation with those workers who are categorized as "good" or "best" or "competent." This shift is credited to the research done by K. Lewin, G. Lippit and R. K. White (Thomas, 1979).

"Teacher Skills and Competencies" Method

Lewin, Lippit & White based their research on skills and competencies. The result was process scales with evaluation instruments that measured the extent to which particular skills surfaced. They emphasized relationships between people and demonstrated skills. Among these were "good classroom climate" and "appropriate rapport with students." Other skills identified by them were "demonstrated abilities to organize," "to prepare adequately," "to inspire," "to develop self-direction in students," "to present clear and definite assignments," "to ask clear and concise questions," "to listen effectively," "to tolerate tension" and "to personalize discipline" (Thomas, 1979).

M. D. Thomas (1979) when writing for Phi Delta Kappa found a checklist of 27 "skills." Even as late as 1975, a San Francisco suburban school system developed and used a long checklist for purposes of evaluating teachers. Among the "skills" the evaluation system listed for rating by the principal were: Lesson is well organized; Careful planning has been done; Presentation is inspiring; Assignments are clear and definite; Harmony exists between teacher and student; Teacher invites differences of opinion; Teacher uses humor when appropriate; and, Teacher is democratic and fair.

Evaluators of teachers used skills, checklists, and competencies to rate teachers. Educational research by M. Mohan, R. E. Hull, B. Rosenshine, and N. Furst correlates the behavior of effective educators with certain competencies (Mohan, & Hull, 1975). They hesitantly theorized that effective teaching is connected to a "cluster" of competencies, rather than to individual skills. Chief among these were "ability to plan effectively," "ability to motivate," "ability to stay on task," "ability to speak clearly," and "ability to control and be respected." Although these competencies are desirable, they are only a part of the performance of teachers. Furthermore, these competencies are difficult to

measure; therefore, the need for another method of evaluation to determine the quality of a teacher appeared. Repeatedly when determining how well teachers do their jobs, these "skills" became questionable. Their appropriateness also became questionable when determining if teachers should be terminated or have their contracts extended (Thomas, 1979).

"Product Evaluation" Method

The next shift in teacher evaluation was to "product evaluation method." This form of evaluation was based on student achievement, test scores, number of students failed, vandalism, attendance, number of graduates who enter college, along with other 'objective' data. The method of product evaluation according to Thomas (1979) was limited by its failure to consider the numerous variables that affect students and their performance. However, this method was said to be superior to the old teacher checklist of traits. Caution regarding the use of the product evaluation method came from Henry Chauncey, former president of Educational Testing Service, "Standardized tests of student achievement are such useful teaching tools that it is often a mistake to try to make them do double duty as measures of the teacher as well." (Thomas, 1979, p. 14).

"Performance Evaluation" Method

Beginning in 1975, some school districts and institutions of higher learning used performance evaluation as a method for evaluating teachers. This new approach incorporated "contributions" made by the educator to the education profession, the community, and the schools and to professional societies and organizations. Therefore, the evaluation of teachers resulted directly from offices held, articles published, speeches given or notes received, honors obtained, and number of times names recorded on conference programs (Thomas, 1979).

"Performance Standards" Method

Performance evaluation of teachers became the preferred practice for evaluating teachers. Its form, however, changed over time. Soon came the replacement for performance evaluation--performance standards. Standards related directly to the job to be done by the teacher-employee in a specific school system. Standards applied to all educational personnel by job classification. Teachers, assistant principals, central office personnel, and superintendents were united by a requirement setting forth the evaluation

of all certificated education employees. Distinction was not made by job or position or degree, taking into account different preparation, roles, and responsibilities within the school system (Thomas, 1979).

The significance of performance standards was their establishment of competencies for all educators from the superintendent to the teacher. Standards did this in a conclusive manner, but more importantly, they set forth desired outcomes, and the accountability system by which all educational personnel would be held (Thomas, 1979).

Performance evaluation was innovative for authentication of achievement. Performance standards could be attained in a variety of ways, but the measurement of performance standards was always the essential final step. Either qualitative or quantitative methods of measurement could confirm the achievement outcomes. Measurement, however, could never be omitted. Without measurement, verification of attainment, or non-attainment of standards, was impossible (Thomas, 1979).

Other studies of teacher evaluation, beyond those of Thomas, revealed past methods of evaluating teachers which were ineffective (Soar, Medley, and Coker, 1983). Specifically, were identified three ineffective methods. First, tests that measured characteristics were identified such as the National Teacher Examination (NTE) and minimum competency tests as ineffective. Second, test scores of students of individual teachers and/or students of groups of teachers on achievement tests such as the Stanford (SAT) were identified as ineffective. Third, ratings of teachers' classroom performance were identified as ineffective.

From the recording of educational history until the 1950s, research on the evaluation of teachers focused on identifying specific teacher characteristics that separated teachers into the two extreme spheres of competence and incompetence (Soar et. al., 1983). Arvil Barr (1979) conducted a wide-reaching review of the literature and concluded that the vast majority of previous studies used supervisors' ratings as the sole measure of teacher effectiveness or competence. Data collected about these rating indicated that teacher evaluations founded on these ratings reflected more about the beliefs of the raters than the extent of individual teacher performance and competency. "Empirical data that exist," state Soar et. al. (1983, p. 245), "indicate that teachers who

are highly rated are no more effective, on the average, in producing student achievement gains than teachers who are rated low."

Some studies on teacher characteristics used average gains in student achievement. A chief criterion was teachers' intelligence. Resulting data were highly variable, but the median correlation between the two factors of effectiveness and ineffectiveness for all of these studies was .03. Research, therefore, does not support teacher characteristics as a viable method for evaluating teacher performance. In fact, none of the studies found a correlation between teachers' scores on any written test and their ability to increase pupil performance (Gage, 1963).

"Student Achievement" Method

Student achievement argue some researchers is the most logical criterion for teacher evaluation. Soar et. al. (1983) suspects this premise for three reasons: student variability, the regression effect, and the limitations of available achievement tests. These researchers further assert that teachers are not analogous to factory workers who work with raw products that are unchangeable, enabling supervisors to predict expected products. Rather, teachers are professionals who interact with changeable human beings of various ages and abilities and conditions, causing superiors of teachers who desire to establish production results for teachers to be proven false by research.

"Merit Pay" Method

The next era of research on teacher evaluation occurred in the 1980s. Its focus was rewarding effective teachers. The preferred method of evaluation was career ladder plans. These incentive payment plans were also known as performance-based incentive programs or differentiated staffing (Hawley, 1985; Edelfelt, 1985). Fueling these programs was the public's desire for teachers to be accountable. Ransom and Knoll were two of twenty-eight states to enact career ladder plans; some state legislatures mandated this requirement.

"Behavorial Standards for Teacher Performance" Model

Based upon the effective schools literature, behavorial standards were studied and codified as behaviors used by teachers whose students performed well. These behaviors generally were classified as competencies in eight categories: 1) Preparation for Instruction, 2) Assessment, 3) Presentation of Instruction, 4) Classroom Management, 5)

Professional Responsibilities, 6) Professional Development, 7) Communication, and 8) Positive Learning Climate (Orientation Manual, 1997).

Each of the eight competencies was subdivided into narrower behaviors named indicators. Indicators were subdivided into smaller, precise behaviors that could be rated. This continuous subdivision removed the high inference level or subjectivity so commonly associated with teacher evaluation. This pilot case study examined this model --the State of Ransom Teacher Evaluation System.

"Peer Review" Method

Since the development of the State of Ransom Teacher Evaluation System, teacher accountability and evaluation has shifted away from Competencies and Standards models. A paradigm shift evolved to contemporary models that are collaborative. Research from the late 1990s, continuing into the twenty-first millennium, reports the use of teacher peer review methods, parents' rating of their children's teachers, students giving their opinions of their teachers, job-specific criteria and a guiding conception of teaching (Bushweller, 1998; MacLean, 1994; Dwyer, 1998). Researchers describe some of these approaches as problematic. However, the commonly used methods for evaluating teacher performance seem to be waning in choice and accuracy. The National Education Association (NEA), one of the nation's two teacher unions, altered its position of opposing teacher peer review. Both NEA and American Federation of Teachers (AFT) now support teachers evaluating other teachers (Bushweller, 1998).

Methodology

As a person involved in the evolution of teacher evaluation models and systems within the State of Ransom and within my local school district, I had a vested interest in understanding how principals within my district perceived the usability of this teacher evaluation system. This pilot case study addresses whether principals who are rigorously and extensively trained would follow the Ransom plan. Further, this pilot case study provided me with information to continue doing research for my doctoral dissertation. Data sources for my research included the historical perspective of teacher evaluation, general survey questionnaires, structured interviews, and experiences of the researcher

when functioning as a trainer of evaluators, school district coordinator of CEPE. and evaluator of secondary teachers.

Participants

Twelve elementary and middle school administrators contributed data to the pilot case study. All of them as required by the Ransom CEPE System obtained certification as evaluators of teachers; no one without this certification could evaluate teachers in Ransom. Of the twelve evaluators, seven were principals and five were assistant principals. Additionally, seven were evaluators at the elementary grades K-5, and three at the middle school grades 6-8. Surveys from administrators at grades 9-12 were not received for this pilot case study. They will be included in my research for my doctoral dissertation.

Protocols

Data were collected using an open-ended survey questionnaire, a structured interview, and reflections of this researcher as a Ransom certified trainer of evaluators, as the CEPE Coordinator for Anderson City Schools, and as evaluator of secondary teachers. The structured interview followed the model of the CEPE structured interview for teachers. Each instrument had a different purpose. The questionnaire examined the experience of all evaluators. The structured interview examined factors that might influence the evaluation process and outcome. Personal reflections examined all aspects of the process and the system from three varied perspectives. All of these used the historical perspective as a framework from which to examine teacher evaluation.

Findings and Discussion

Section I of the open-ended survey questionnaire attempted to discover demographic data of participants through their answering of ten questions. Survey responses to questions one through four addressed position, level, category of teachers evaluated, and number of teachers evaluated by each evaluator. Responses to question one showed that the twelve certified evaluators were seven principals and five assistant principals. Survey responses to question number two showed that the majority was elementary followed by middle school administrators.

Question two from Section I of the open-ended questionnaire sought the category of teachers evaluated by each evaluator. Choices were A) First Year Only; B) Non-tenured (second, third year) Only; C) Tenured (four, five, and more years) Only; D) First year and Non-tenured (first, second, third year) Only; E) Non-tenured and Tenured Only; and F) First year and Tenured Only. Responses to question two revealed that seven of twelve respondents evaluated a combination of non-tenured and tenured teachers. Specifically, four of twelve respondents evaluated only non-tenured teachers who are defined as teachers ranging in experience from first year to second year to third year within the school system. In contrast, tenured teachers are defined as teachers who range in experience upward of three years and who have been awarded tenured status by vote of the board of education. Finally, findings in response to question two showed that one of twelve participants evaluated tenured teachers only. Thus, eleven respondents or 92 % of all evaluators placed a combined classification of tenured and non-tenured teachers into the evaluation cycle. One respondent evaluated only tenured teachers.

Next, question three sought the number of teachers evaluated by each evaluator. Twelve participants evaluated from five to eight teachers. This represented 100 % of the participants. None of the participants chose to evaluate from one to four teachers nor from nine to twelve teachers nor from thirteen to sixteen teachers.

Responses to questions number two and three suggested that a sizable percentage of teachers on each faculty are non-tenured. Evaluators almost without exception evaluated only non-tenured teachers. The selection of non-tenured teachers also suggested that evaluators considered their seemingly unending responsibilities in relationship to their limited time for evaluation. Therefore, it appears that they too-like their predecessors of past decades-looked for the most direct way to evaluate teachers. Their predecessors used checklists; they focused only on those teachers who by law and policy must be evaluated annually.

Section II of the open-ended questionnaire centered on opinions and experiences of the evaluators. When asked to rate the training they received for CEPE, eleven respondents rated the training as "Just Right." These evaluators rejected choices of "Too Much" and "Too Little." In total, ninety-two % of respondents marked the effectiveness

of their CEPE training as "Just Right" while one respondent considered the effectiveness of their training as "Too Little."

Next, respondents indicated their need for refresher training. Although choices of areas for refresher training were numerous and comprehensive, they were overwhelmingly discarded by respondents. Categories of "Procedures," "Timelines," "Holistic Scoring," and "Other" received one request each for refresher training. The topic of Professional Development was the lone write-in choice for the "Other" category. Eight participants created and wrote in the category of "NONE." This represented 67 % of the participants. The next question sought evaluators' views on the ease or cumbersomeness of evaluation instruments or forms. Given the choice of selecting the **two** most cumbersome instruments, participants gave diverse ratings. Yet, a clear majority, 75 %, agreed upon one choice. Their first instrument of choice was the scripting instrument for Classroom Observation. The second most cumbersome instrument identified by them was the scripting instrument for the Structured Interview. Thirty-four percent chose this option. The Structured Interview instrument and the Classroom Observation Analysis instrument each received the most cumbersome ratings of twenty-five % of the participants. The final choices were the Supervisor's Review Form and the Evaluation Summary Report. Both of them received a rating of 8 %. Administrator-practitioners often recount that many of their responsibilities lead to interruptions which tend to interfere with duties like evaluation of teachers. Administrators in this pilot case study were asked to identify all obstacles they had to overcome to implement CEPE in a timely manner. Offered as choices were nine possible obstacles: A) student discipline, B) personnel disharmony, C) suspensions/expulsions, D) parent conferences, E) unannounced visitors, F) telephone calls, G) school system obligations, H) fund-raisers, J) athletics & clubs and K) others [to be listed]. Study participants identified all obstacles faced by them. A breakdown of results follows. "Student discipline and school system obligations" tied for first place with 42 %. "Parent conferences" ranked second with 34 % of the respondents. "Suspension/expulsions" and "telephone calls" with 25 % shared third place ranking. Fourth was "unannounced visitor" with 17 %. "Fund-raisers" had only 8 %.

The purpose or goal of the Ransom CEPE System is to improve teaching and learning in Ransom schools. To determine the level of achievement for this goal, the open-ended survey requested evaluators to rate the status of effective teaching in their respective schools. They were also asked to rate the status of effective learning in their respective schools. On a scale of "A," "B," "C," "D," to "E" or low to high rankings, evaluators reported the quality of effective teaching practiced in their schools. Eight selected "D" for a rating of 67 %. Category "C" received three marks for a 25 % rating. Category "B" received a rating of 8 %. Categories "A" {lowest} and "E" {highest} did not receive any markings. Likewise, using the same scale as outlined before, respondents reported the quality of effective learning in their respective schools. Results were similar between the two groupings of teaching and learning. Again, categories "A" {low} and "E" {high} did not receive any markings. Category "B" received one response, an 8% rating. Category "C" received four markings for a rating of 34 %. Finally, Category "D" had the top rating of seven markings for a rating of 58 %.

The CEPE System also prescribed the daily instructional design or format for lessons taught in Ransom schools. Classroom observations rated the use of this format titled **"The Flow of A Lesson."** The researcher's aim was to determine the degree of implementation of the state plan in classrooms. The question, "Indicate the percentage (%) of teachers in your school who use FLOW of A LESSON for daily instruction" was asked. Eleven responses, in increments of five from 00 to 100, were possible; yet, only six selections were chosen. Four (4) evaluators marked 95% use; three evaluators marked 85% use; and two evaluators marked 55 % use. Remaining percentages used were 25 % and 75%. None of the respondents marked 100 % use of the CEPE instructional design **"The Flow of A Lesson."**

Another data source for this pilot case study was the structured interview. Its purpose was to examine potential influencing factors such as gender, race, evaluator-teacher relationship, socio-economic status of students, location or neighborhood of school. Participants for the structured interview were selected and individually invited to participate by the researcher. Potential participants represented the school system's diversity. Just as for the CEPE system, the structured interview was developed and presented to participants before requesting their participation. None of the selected

participants declined to participate; however, scheduling conflicts at the secondary level prevented their participation. Thus, as the researcher, I conducted three structured interviews.

The school system's diversity included different grade levels or divisions, race, gender, enrollment, and level of participation in the federal free and reduced breakfast and lunch program. Participants were from two elementary schools and one middle school. Two participants were black; one was white. Two were male; one was female. One school had a high free lunch and breakfast enrollment; two schools had a mid-level free lunch and breakfast enrollment.

Question Set One asked, "How do you rate CEPE? Does CEPE correctly evaluate the performance of elementary teachers? Does CEPE correctly evaluate the performance of middle and secondary teachers? Does CEPE benefit teachers of any particular content areas?" When responding to Question Set One, which seeks to ascertain if CEPE meets its goal of accurately measuring teacher performance, principal evaluators gave CEPE very high ratings. One respondent, said, "In my seven to ten years as an administrator, CEPE is the best evaluation system I have used." A second respondent called it, "Excellent! It helps the teacher. It helps the administrator. I like it when we are on the same page together. That's the only way the school can improve. Everyone understands that we have to help students achieve. Everyone knows also how we can do it." A third respondent thought CEPE was complete. "I did not observe any differences in it when I observed the various grade levels and subjects taught. In fact, it is so structured that I wish the other systems (counselor, librarian, special education teacher of students where academics is not the major focus) had been developed and implemented this year."

Question Set Two asked, "How do the teachers you evaluated with the CEPE System perceive you as an evaluator? Do the teachers you evaluated this year perceive you as an evaluator differently than an evaluator with any other evaluation system such as Knowledge? Did factors such as sex, race, and/or school neighborhood contribute to teacher perceptions of you as an evaluator?" This question set dealt with the perceptions of teachers about the evaluator as each functioned in the role of CEPE evaluator, of principal, and of human being. Without hesitation, all three respondents said teachers, to their knowledge, perceived them as fair. Nothing in their experiences caused them to

believe that teachers had brought outside factors such as race, gender, and location, or neighborhood of school, to the evaluation process.

Question Set Three asked, "How accurately did Observation Two and/or Observation Three assess teacher delivery? Elementary and Middle Levels: Consider that any observation of less than 45 minutes had to be voided and begun again for 45 minutes. An emergency is not an exception. Secondary Level: Consider that the evaluator observed 45 minutes of the middle of a class for Observation Two and 45 minutes of the end of a class for Observation Three, then rated the teacher on the total lesson for the entire period according to '**Flow of A Lesson**.'" Question Set Three also asked, "How accurately did the structured interview assess teacher planning? What was the result of using the oral option? What was the result of using the written option? Was there any significance in scoring between the two options?" Question Set Three focused on the second and third classroom observations as well as the structured interview. Did variations between regular and block scheduling matter? Did selecting the written vs the oral option of the structured interview benefit or harm the teacher?

Evaluators similarly reported the lack of difference between the written and the oral options of the structured interview. Principals planned with their respective secretaries the avoidance of interrupted observations. They planned to observe for the full required 45 minutes; only a life-threatening situation would have caused them to stop, necessitating their return to observe for a full 45 minutes. They were always aware that any shortages of time less than 45 minutes would void the observation.

Because there were not any participants from the secondary level, findings regarding block scheduling had to be moved to my doctoral dissertation for study. I project, however, that practices and experiences will vary between the secondary level and the lower levels. This projection is based upon my personal reflections. They are the final data source for this pilot case study.

This pilot case study attempted to examine the implementation of CEPE for teachers of grades pre-kindergarten to twelve. However, participation by secondary administrators was not obtained. The result is that the secondary experience can only be collected through one source: The researcher's varied and overlapping roles as Ransom

certified trainer of evaluators, as CEPE coordinator for my school system, and as evaluator of teachers for grades 9-12.

Training of trainers of evaluators was a high priority of Ransom's CEPE. Personnel to be trained as a trainer of evaluators were selected by state officials and superintendents. Our training sessions were highly charged with competition. No one wanted not to qualify. After all, district pride and obligation weighed heavily as much as, if not more than, personal pride and obligation. CEPE officials at the state level stressed repeatedly that CEPE for teachers would only be as effective as we trained the potential evaluators-principals and assistants. We potential trainers went to training sessions armed with the knowledge that someone would not qualify. Therefore, we went early - 8 A. M. - and stayed late – 7 to 8 P. M. We studied hard, skipping meals and other pleasures if deemed necessary. What a relief when I passed the three required reliability tests in modified scripting and the knowledge test at more than 80% mastery. In fact, my raw score was 92. I could return to Anderson with pride and a sense of fulfillment. So did most of the other potential trainers. We suspected that indeed several who did not qualify would not be known. Our trainers were so professional that the group could only assume who might not have passed.

Training for CEPE: The Teacher System was the third round of training for me. Previously, I trained and became a RANSOM certified evaluator of the administrator system and the first teacher system. My rigorous training the third time around caused me to be overzealous in preparing for and delivering training to potential evaluators. It was my pleasure to train evaluators in three districts; I received rave reviews and made some lifetime friends and colleagues. It appears that I converted what had been a grueling training experience for me into an ardent, but pleasant experience for others.

My school system has two central office administrators and one assistant principal who are Ransom certified trainers of evaluators. It was the intent of the Ransom State Department of Education and the local superintendent and school board that local trainers would train other local administrators as CEPE evaluators for teachers. This was how the State of Ransom intended to facilitate the implementation of CEPE for teachers at the local level. However, in some districts, including mine, this procedure was not followed. Local administrators from around the State chose to be trained by trainers outside of the

local school system. Local trainers, however, often had the role of assisting the outside trainer. It appeared that local administrators were concerned about their future relationship with local trainers after CEPE training had ended. Some local administrators were fearful of the outcome just as the potential trainers had been fearful. Daily headlines in newspapers throughout Ransom fueled the fear as they recounted tales and predictions of principals and other administrators who could not pass tests themselves. How could they fulfill their job responsibilities if they, themselves, could not pass high-stakes tests, but still expected teachers and students to pass high-stakes tests?

My role as CEPE coordinator was essential to my school system's understanding and implementation of the Teacher System during its Start-Up Year. During the course of the year, I provided evaluators with technical assistance. I also guided evaluators through school responsibilities such as the required orientation session before any implementation of CEPE for teachers. When requested, I conducted orientation sessions. Many times evaluators had procedural questions. They never forgot that teachers could not disagree with ratings. Appeals could only result from a violation of procedures. Therefore, my chief role was to provide information and give reminders of timelines. Thus, cooperation was very high. It seemed that almost everyone was committed to implementing this new evaluation system precisely as prescribed by the Ransom State Department of Education.

At the secondary level, grades 9-12, I was the evaluator of eight teachers. They taught language or arts education or career-technical subjects or disabled students. Six were female; two were male. All were non-tenured teachers, but one was a first year teacher pursuing one of the newly developed alternate ways to certification.

The Ransom State Department of Education wished to assure teachers that the evaluation system was not "out to get them." The evaluation system was enacted to improve performance; the total system pivoted upon the professional development component. Tenured teachers had to be provided with professional development opportunities for learning and implementing new skills before termination of contract became an option. Non-tenured teachers also had to be provided with professional development opportunities. However, continuation of their contracts was determined by personnel laws. To effect the assurance that the Ransom Department of Education wished to give to all teachers, the Department of Education stipulated that **EVERY** teacher

would be provided with a personal copy of the Orientation Manual. Its content paralleled the Evaluator Manual. Evaluator and evaluatee (teacher) had the same information, procedures and instruments. Training was the only difference in their method of operation and preparation.

Before beginning the evaluation cycle, I, as evaluator, presented myself to each teacher for whom I would be their evaluator for the year. I went on to offer the answering of any questions they might have. I ascertained their possession of a personal copy of the Orientation Manual. I also recommended that they listen attentively at the upcoming required Orientation Session. If they did not know about the second upcoming orientation session that was to be conducted by Ransom Education Association, the state teachers' union, I made them aware and advised them to attend. I further advised them to read closely all portions of the Orientation Manual. Before time for each instrument, I recommended that they would examine appropriate terms and tips. I operated here from the knowledge that change is difficult. Familiarity with the unknown reduces apprehension. I also knew that the more the teachers knew about "the what" and "the why" and "the how" of each stage of CEPE, the more responsibility they would assume for their evaluation outcome and professional development.

My analysis of these occurrences is that those teachers who did their homework had little difficulty with CEPE. Their attitudes were positive; their energy levels were high. They began to assess their teaching behaviors, and sought methods of improvement. They thought outside of the box when dealing with problematic situations. Their emphasis was on effectiveness.

Where problems occurred, teachers were operating out of another evaluation paradigm. They operated outside of the school system's paradigm. They appeared unwilling to change as evidenced by their behavior. Even when informed of CEPE's "Look Fors" and of the strong possibility that I would return within one week, I did not see any practices that had been altered from previous visits.

Teachers who desired higher ratings suggested that I just did not know where to look. When presented with supporting evidence taken from the scripts made during the observations and the structured interview conferences, I was informed that I didn't understand what was happening that day. Perhaps if I came at more regular times, I

could see what was actually happening. Nevertheless, at any rate, could it be that my mind was already made up?

All of the teachers with positive attitudes were young. Many of them were female; however, one male teacher matched them in enthusiasm and follow-up. A shift in instructional and evaluation paradigms would benefit some teachers of both sexes. Preservice education appears to be preparing future teachers for the world of today and yesterday and tomorrow. In turn, they can prepare their students for the unknown adventures of the new millennium.

Conclusions and Recommendations

The Anderson Schools Experience

The state-mandated CEPE evaluation system has been successfully delivered to the Anderson School System. Evaluators demonstrated consistency in implementation. Their training was adequate. Students and teachers advanced in learning and teaching because of the CEPE system. CEPE was the standard bearer. It educated and evaluated, which has caused school improvement. That is the ultimate goal of any evaluation system.

Leadership styles quietly influenced daily and periodic occurrences. Positive, knowledgeable, evaluators took their apprehensive teachers by the hand and led them through the CEPE maze. Their individual leadership styles were the most critical factors throughout the pilot case study. Despite their previous and extensive training, evaluators accepted their tasks and became CEPE ambassadors. They fulfilled requirements and lauded CEPE as the best system yet.

At the end of the year, evaluators offered recommendations for improving the implementation of CEPE. One recommendation coming from their experience was the elimination of a third observation for teachers with one year or more of experience. Second, concurrence with three observations for beginning or first year teachers was strong. Additionally, evaluators acknowledged the vast amount of time needed to use CEPE, and wished there were some way to reduce the huge blocks of needed time. Nevertheless, evaluators acknowledged the benefits of CEPE. They liked CEPE and anticipated further improvement in teaching and learning because of it.

Overall

The Anderson experience as examined in this pilot case study suggested that CEPE met the needs of all school stakeholders in the State of Ransom. Teacher evaluation or accountability was accurately assessed, measured and used for professional growth. School improvement was the victor.

The CEPE: Teacher System is an effective system (as defined by educational literature) that accurately measures teacher performance. Extensive training for all evaluators ensured the design of the system transferred from the Ransom State Department of Education to singular classrooms across Ransom. Evaluators turned theory into practice and strengthened schools. Attention to preciseness, organizational skills, and extensive knowledge as well as comprehension was the responsibility of each evaluator. Their acceptance of this responsibility resulted in the implementation of CEPE: The Teacher System to be credible and effective.

In summary, the contradictions within the state-mandated system were several: First, from what educational research says about change, mandating any system is not viewed as the most effective means for change. Yet another contradiction appears to be that the system holds promise for assisting administrators with a useful tool for improving teacher performance and student achievement. A third contradiction was that the teacher evaluation system was presented by the media as a punitive system to get rid of "bad" teachers. The truth was that in fact the system had the dual purpose of improving teacher performance and providing professional development in areas of need. A fourth contradiction is the accuracy and objectivity of results that were obtained only by collecting, analyzing, and compiling data for a minimum of eight hours per teacher. There has to be some simple method that accurately measures teacher performance yet does not consume time. However, as this pilot case study shows the needed simple method is neither checklists nor any other method used in the past to evaluate teachers. The larger contradiction may be in districts and states that expect administrators to be "paper pushers" rather than focusing in on teacher improvement.

The researcher acknowledges that the research base for the studied teacher evaluation system is not the most up to date. However, CEPE: The Teacher Evaluation System appears to have made major improvements in teacher performance, student

achievement and school reform in the State of Ransom. Because of it, some systems now evaluate teacher performance; they did not have any system in place before CEPE. Because of it, some systems now assess their effectiveness in contrast to the past when they simply did what they had always done. Finally, because of it, some systems now have a tool by which they can identify and advance improvements in staffing, curriculum, technology, instructional purchasing, counseling and library needs, student learning, student achievement, student development, parental and community involvement as well as school facilities. As teachers improve, students and schools will improve. Students of Ransom deserve no less; they are the visionaries, builders and keepers of the future.

References

Alabama State Department of Education. (1997). Alabama Professional Education Personnel Evaluation Program For Teachers: Evaluator Manual, Orientation Manual, Trainer Manual (April 1, 1997). Montgomery.

Barber, L. W. & Klein, K. (1983). Merit pay and teacher evaluation. Phi Delta Kappan, 65 (4), 247-251.

Boyd, R. T. C. (1997) Improving teacher evaluation. ERIC Digest, [On-line], 111. Available: America Online.

Bushweller, K. (1998). Other voices: Listening to what fellow teachers, parents, and students have to say in teacher evaluation. The American School Board Journal, 185 (9), 24-27.

Conley, D. T. & Dixon, K. M.(1990). The evaluation report: A tool for teacher growth. NASSP Bulletin, 74 (527), 7-14.

DePasquale, D. Jr. (1990). Evaluating tenured teachers: A practical approach. NASSP Bulletin, 74 (527), 7-14.

Dwyer, C. A. (1998). Psychometrics of praxis III: Classroom performance assessments. Journal of Personnel Evaluation in Education, 12 (2), 163-187.

Edelfelt, R. A. (1985). Career ladders: Then and now. Educational Leadership, 43 (3), 62-66.

Gage, N. L., ed. (1963). Handbook of research on teaching. Chicago: Rand-McNally.

Gitlin, A. & Smyth, J. (1990). Educative possibilities in teacher evaluation: Two alternatives. NASSP Bulletin, 74 (527), 25-34.

Hawley, W. D. (1985). Designing and implementing performance-based career ladder plans. Educational Leadership, 43 (3), 57- 61.

House, E. R., Rivers, W. & Stufflebeam, D. L (1974). An assessment of the Michigan accountability system. Phi Delta Kappan , LV (10), 663- 669.

Isenberg, A. P (1990). Evaluating teachers—Some questions and some considerations. NASSP Bulletin, 74 (527), 16-18.

Katims, D. S., Gilbreath-Coleman, K. K. & Jones, M. S. (1989). Teachers perceptions of the Texas appraisal system. Teacher education & practice (2), 25-32.

Keefover, K. S. (1983). Accountability-a historical perspective. The Educational Forum, XLVII (3), 365-372.

Kher-Durlabhji, N., Lacina-Gifford, L. J. & Hunt, S. N. (1992). Louisiana teacher evaluation program (LaTEP): Teachers' perceptions. Record in Educational Administration and Supervision, 12, (2), 22-27.

Koehler, M. (1990). Self-assessment in the evaluation process. NASSP Bulletin, 74 (527), 40-45.

Lofton, G. G., Hill, F., & Claudet, J. G. (1997). Can state-mandated teacher fulfill the promise of school improvement? Events in the life of one school. Journal of Personnel Evaluation in Education, 11 (2), 139-165.

MacLean, J. (1994). Personnel evaluation in education: A model for job-specific criteria. The Clearing House, 68 (1), 47-51.

Manning, R. C. (1988). The teacher evaluation handbook. Englewood Cliffs, NJ: Prentice Hall.

Pope, C. A. (1990). Indirect teaching and assessment: Are they mutually exclusive? . NASSP Bulletin, 74 (527), 1-5.

Root, D. & Overly, D. (1990). Successful teacher evaluation—Key elements for success. NASSP Bulletin, 74 (527), 34-39.

Sergiovanni, T. J.. (1977). Reforming teacher evaluation: Naturalistic alternatives. Educational Leadership, (64), 602-607.

Soar, R. S., Medley, D. M. & Coker, H. (1983). Teacher evaluation: A critique of currently used methods. Phi Delta Kappan, 65 (4), 239-246.

Texas Education Agency (1988). Texas teacher appraisal system teacher orientation manual. Austin, Texas: Division of Management Assistance and Personnel Development.

Thomas, M. D. (1979). Performance evaluation of educational personnel (Fastback). Bloomington, IN: Phi Delta Kappa Educational Foundation.

Wareing, C. (1990). Up close and personal: A model for supervision and evaluation. The Clearing House, 63 (6), 245-250.